The Patent Office

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.
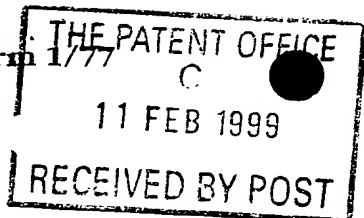
PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Signed

Dated   23 February 2000

An Executive Agency of the Department of Trade and Industry

THE PATENT OFFICE
C
11 FEB 1999
RECEIVED BY POST

The
Patent
Office

1/77

11FEB99 E424349-1 002934
F01/7700 0.00 - 9902970.4

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1.	Your reference	PHM 99-001

2.	Patent application number *(The Patent Office will fill in this part)*	**9902970.4**

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

Zeneca Limited
15 Stanhope Gate
LONDON
W1Y 6LN, GB

Patents ADP number *(if you know it)*

6254007002

If the applicant is a corporate body, give the country/state of its incorporation

4. Title of the invention

NOVEL MATRIX

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

BILL, Kevin

ZENECA Pharmaceuticals
Intellectual Property Department
Mereside, Alderley Park,
Macclesfield, Cheshire, Sk10 4TG, GB

Patents ADP number *(if you know it)*

4469847002

6.	If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number	Country	Priority application number *(if you know it)*
			Date of filing *(day / month / year)*

7.	If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
   a) *any applicant named in part 3 is not an inventor, or*
   b) *there is an inventor who is not named as an applicant, or*
   c) *any named applicant is a corporate body. See note (d))*

9. Enter the number of sheets for any of the
   following items you are filing with this form.
   Do not count copies of the same document

   Continuation sheets of this form

   Description   46

   Claim(s)

   Abstract

   Drawing(s)

10. If you are also filing any of the following,
    state how many against each item.

    Priority documents

    Translations of priority documents

    Statement of inventorship and right
    to grant of a patent *(Patents Form 7/77)*

    Request for preliminary examination
    and search *(Patents Form 9/77)*

    Request for substantive examination
    (Patents Form 10/77)

    Any other documents
    (please specify)

11.   I/We request the grant of a patent on the basis of this application.

                                         Signature *Lynda M Slack*   Date 10 February 1999

12. Name and daytime telephone number of   Mrs Lynda May Slack   01625 516173
    person to contact in the United Kingdom

Warning
After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes
a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

NOVEL MATRIX

The present invention is directed to an oligonucleotide array ('generic' array) consisting of a plurality of different oligonucleotides of predetermined sequence, attached to a

5 solid surface at predetermined positionally distinct locations, characterised in that the oligonucleotides have substantially the same melting temperature (T_m). The invention is also directed to novel probe and primer targeting polynucleotides which possess nucleotide sequences that are complementary to different target nucleic acids from a test sample as well as to the oligonucleotides on the array. The 'generic' array and targeting oligonucleotides

10 prove useful in a number of hybridisation capture methods and assays such as, sequence identification by allele specific hybridisation, fingerprinting and genome typing, differential gene expression profile analysis, or in clinical diagnostic methods utilising for example, ARMS amplification primers. The present invention is therefore also directed to methods or assays utilising such 'generic' arrays and, in particular, to the use of a multiplex amplification

15 assay that utilises a plurality of ARMS primers possessing non-amplifiable tails in conjunction with a 'generic' array for detection amplification of variant nucleic acid sequences.

Microarray (also termed hybridisation array, gene array or gene chip) technology wherein nucleic acid molecules attached to solid substrates at predefined locations in small

20 areas and at high density are used, in conjunction with hybridisation reactions, for identifying and discriminating target nucleic acid sequences, has advanced rapidly in the past few years. These chips or microarrays allow massive parallel data acquisition and are used, for example, in polymorphism detection, clinical mutation detection, expression monitoring, fingerprinting and sequencing.

25 A variety of methods are currently available for making arrays of biological molecules. The 'dot or slot blot' approach, whereby an ordered array of DNA is vacuum blotted using a manifold, or hand blotted by capillary action, onto a porous membrane, such as nylon or nitrocellulose has been around for many years (Maniatis et al., Molecular Cloning-A Laboratory Manual, First Edition, Cold Spring Harbor, 1982). Methods for preparing a

30 plurality of oligonucleotide sequences and for attaching these to solid supports at high density are also known in the art. For example, US Patent No. 4,562,157 describes a method of using

photo-activatable cross-linking groups to immobilise pre-synthesised ligands on surfaces.
Fodor et al. (Nature. 364:555-556, 1993) and US Patent No. 5,143,854 describe the 'light-
directed chemical synthesis' method for synthesising ligands, including oligonucleotides,
directly onto a substrate surface at the desired location. US 5,700,637 also describes methods

5 for *in situ* synthesis of oligonucleotides on solid support surfaces. In addition, such methods
for preparing microarrays can easily be automated. International Publication No. WO
95/35505 discloses an automated capillary dispensing device and method for applying
biological macromolecules to solid supports. International Publication No. WO 97/44134
also describes devices for delivery of small volumes of liquid (which may contain biological

10 macromolecules) in a precise manner to produce microsized spots on a solid surface to
generate a microarray. Similarly, International Publication No. WO 98/10858 also describes
an apparatus for the automated synthesis of molecular arrays. Techniques exist for applying
the oligonucleotides to the array at high density and for example, techniques exist for
applying well in excess of 10^3 distinct polynucleotides per 1 cm^2.

15   Many of the advances in microarray technology concern increasing miniaturisation.
Aside from the ease in handling and manipulating smaller hybridisation matrices, one
significant advantage that smaller chips with higher density of capture probe have over larger
formats is that the sample does not have to be "stretched out". The technology is unlikely to
be widely accepted in the clinical diagnostic market however until costs have been

20 substantially reduced from their current levels.

        Oligonucleotide DNA arrays consisting of short oligonucleotides (e.g. typically 8mers
to 20 mers) bound via their 3' termini to a solid surface such as glass or a silicon wafer have
been proposed as tools for mutation detection or for the resequencing of genes (e.g. Chee et.
al., Science, 1996, Vol. 274, pp 610 - 614; Drobyshev et. al., Gene, 1997, vol.188, pp 45-52).

25 The mechanism of analysing DNA sequences depends on the principles of allele specific
hybridisation. In brief, an oligonucleotide array is prepared with a set of overlapping
oligonucleotide probes (e.g. 20 mers) complementary to the consensus sequence of the DNA
target designed so that each sequence is offset from the previous sequence by one base pair.
As well as the consensus sequence all three variant sequences at the central nucleotide

30 position are also included on the array. Often the probe sequences for analysing the opposite
template strand are also included on the array. The target sequence is amplified using the

polymerase chain reaction (PCR) and the products are then often transcribed into RNA which has incorporated therein a fluorescent label such as fluorescein-UTP. The labelled RNA is sheared into short fragments and then hybridised to the DNA array. The most stable hybrids form when the target sequence binds to its fully complementary sequence on the array. If

5   there is a mutation in the target sequence then it will bind to one of the variant probes more effectively than to the consensus probe. The relative extent of binding of the target sequence to the probes is measured by monitoring the intensity of fluorescence at each site on the DNA array. If the target sequence is mutated then the fluorescence intensity will be greater at one of the variant probe positions than at the consensus probe position.

10   Microarray technology also makes it possible to simultaneously study the expression of many thousands of genes in a single experiment. Differential expression profiles from, for example, normal versus diseased tissues or induced versus un-induced tissues can be obtained by hybridising the product of expressed mRNA to complementary nucleic acid at pre-defined locations on the array. Alternatively, a time-course of expression of thousands of genes over

15   several experiments from a single sample could be performed. Analysis of gene expression in human tissue (i.e. biopsy tissue) can assist in the diagnosis and prognosis of disease and the evaluation of risk for disease. A comparison of levels of expression of various genes from patients with defined pathological disease conditions with normal patients enables an expression profile, characteristic of disease, to be created. There are currently two approaches

20   to analyse gene expression using microarrays. In the first approach, cDNA fragments, often generated by PCR, for each of the genes under study are attached to an array. Typically, mRNA isolated from the test samples (i.e. induced or un-induced) is reverse transcribed into cDNA with incorporation of a fluorescent label. The cDNA is sheared and hybridised to the array. The other test sample mRNA can be reverse transcribed with incorporation of a

25   different fluorescent label to enable direct comparison of the expression level of each test gene on the same array (see WO 95/35505). The second approach is similar to the first except that an oligonucleotide microarray is used. Because of the differences in hybridisation properties between short oligonucleotide probes, each gene must be represented by several oligonucleotides (typically 20 or more) on the chip. In addition, a partner control

30   oligonucleotide identical to each oligonucleotide, except for one of the central nucleotides, is included on the array to serve as an internal control for hybridisation sensitivity. Thus,

whereas cDNA arrays only require each gene to be represented by a single hybridisation partner on the array, with oligonucleotide arrays, each test gene must be represented by approximately 40 distinct oligonucleotides each at a different position on the array. The advantage of oligonucleotide arrays over cDNA arrays however, concerns the shelf-life of the

5 sample on the array. In general, a cDNA library prepared on an array is useable for weeks whereas, pre-prepared oligonucleotide arrays can be stored for considerably longer.

The strengths of the DNA microarray concept is its ability to carry out very large numbers of hybridisation based analyses simultaneously. However, as the capture sequences attached to the support (chip) have to complement the target sequence, knowledge of the

10 target sequence is required. Each chip has to be custom built on the basis of this known sequence. The need to develop a new custom chip for each new test renders the technology costly and complex. Other concerns involve the hybridisation conditions that must be adopted for each test. Secondary and tertiary structure formation can interfere with hybridisation of the capture and target molecules. In addition, duplex formation between different individual

15 pairs of target sequence and capture sequence may have different stabilities (melting temperatures), because of different G-C content, for example. Current approaches to overcome some of these hybridisation problems include: applying parallel hybridisation across the array, altering the concentration of capture nucleic acid at a particular location, modifying the length of the oligonucleotide at a particular location so as to alter duplex

20 stability, and using tuned electric fields as demonstrated by Edman et al, (Nucleic Acids Research. 25(24):4907-4914, 1997. In practice, however, as DNA duplexes between different individual sequences on the arrays and their cognate complementary target sequences have various different stabilities, custom hybridisation conditions have to be employed for each particular test. In view of the different hybridisation stabilities however, hybridisation

25 conditions adopted are generally not optimal for each and every capture sequence on the array, but are generally a compromise. It would be advantageous to design a microarray wherein substantially the same hybridisation between each pair of target and capture molecule occurs under any chosen hybridisation conditions.

According to a first aspect of the present invention there is provided a solid support

30 having immobilised thereon a plurality of pre-selected oligonucleotides at pre-defined

positionally distinct sites, characterised in that the sequence of each oligonucleotide when bound to its complementary sequence has substantially the same melting temperature (T_m).

With regard to the meaning of "substantially the same melting temperature (T_m).", each single-stranded oligonucleotide immobilised on the solid support has, in increasing order

5   of preference, a T_m when bound to its complementary sequence, within 8°C, 7°C, 6°C, 5°C, 4°C, 3°C, 2°C, 1°C, and 0.5°C of the average T_m of all the oligonucleotides immobilised on the solid support. In a more preferred embodiment the oligonucleotides immobilised on the solid support will possess melting temperatures within a range of 0 to 8°C of each other. In an even more preferred embodiment the oligonucleotides immobilised on the solid support

10   will possess melting temperatures within a range of 0 to 2°C of each other. In the most preferred embodiment, 90-100% of all the oligonucleotides immobilised on the solid support will possess the same melting temperature, and the remaining oligonucleotides will preferably possess melting temperatures within a range of 0 to 2°C of this mode value.

Although it is preferred that all of the oligonucleotides on the solid support fall within

15   the ranges or values for melting temperature as defined above, it is envisaged that a small number of oligonucleotides preferably less than 1-5%, more preferably less than 2% of the total number of oligonucleotides may fall outside these ranges or values.

The melting temperature (T_m) referred to herein, is defined as the temperature at which duplex DNA exists in a ratio of 50:50 in hybridised and dissociated form under

20   equilibrium conditions. The principal governing factors determining T_m are sequence length and G-C content. The theoretical and experimental procedure for determining the T_m is disclosed in Molecular Cloning-A Laboratory Manual, Second Edition, J Sambrook et al., Cold Spring Harbor, Chapter 11 section 46 and 55. In essence, for oligonucleotides shorter than 18 nucleotides, the T_m of the hybrid is estimated by multiplying the number of A + T

25   residues in the hybrid by 2°C and the number of G + C residues by 4°C and adding the two together. For oligonucleotides between approximately 14 and 70 nucleotides in length, the following equation devised by Bolton and McCarthy, (P.N.A.S. **48**:1390, 1962) for determining T_m of long DNA molecules is also applicable:

$$T_m = 81.5 - 16.6(\log_{10}[Na^+]) + 0.41(\% \ G + C) - (600/N).$$

30   Wherein N = chain length and [Na^+] is the ionic strength of the hybridisation solution.

The term "nucleotide" as used herein can refer to nucleotides present in either DNA or RNA and thus includes nucleotides which incorporate adenine, cytosine, guanine, thymine and uracil as base, the sugar moiety being deoxyribose or ribose. It will be appreciated however that other modified bases capable of base pairing with one of the conventional bases,

5 adenine, cytosine, guanine, thymine and uracil, may be used in the oligonucleotides, probes or primers employed in the present invention. Such modified bases include for example 8-azaguanine and hypoxanthine.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more nucleotides (i.e. deoxyribonucleotides or ribonucleotides), preferably more than five.

10 Its exact size will depend on many factors, such as the reaction temperature, salt concentration, the presence of denaturants such as formamide, and the degree of complementarity with the sequence to which the oligonucleotide is intended to hybridise.

In operation, under any hybridisation conditions adopted, all of the oligonucleotides on the solid support (herein referred to as "capture oligonucleotides") that are to be used for

15 capture of a target sequence, exhibit approximately the same hybridisation stability with their cognate complementary sequence as the other pairs of oligonucleotide and complementary target sequence. This ensures that approximately equivalent amounts of target DNA are bound to the complementary oligonucleotide on the array at any particular time, facilitating quantitative analysis. As the T_m of all the duplexed oligonucleotides will be substantially the

20 same however, the optimum temperature for hybridisation can be adopted. With oligonucleotide hybridisation, the optimum hybridisation temperature is generally carried out under conditions that are 5-10°C below the T_m, with the hybridisation and subsequent washes carried out under stringent conditions. Ideally, the hybridisation temperature is controlled precisely, preferably to ±2°C , more preferably to ±0.5°C or better, particularly when the

25 hybridisable length of the capture oligonucleotides are small and there is a need to discriminate between two sequences that may only differ by a single nucleotide at one or other of the termini of the hybridisable sequence. It will be apparent that the capture oligonucleotides need not all be of the same length. Depending on their relative G-C content, two oligonucleotides of different lengths may nevertheless, have the same T_m. Capture

30 oligonucleotides of substantially the same length and G-C content are preferred however.

According to a further aspect of the invention there is provided a solid support having immobilised thereon a plurality of pre-selected oligonucleotides at pre-defined sites, characterised in that the capture portion of all of the oligonucleotides are of substantially the same length and they all have substantially the same G-C content.

5   With regard to the meaning of "substantially the same length", in increasing order of preference, the length of each oligonucleotide immobilised on the solid support will be within or equal to16, 12, 10, 8, 6, 5, 4, 3, 2, and 1 nucleotide(s) of the average length of all the oligonucleotides immobilised on the solid support. In a preferred embodiment the oligonucleotides will each be of a length that is within 0-8 nucleotides of each other.

10   With regard to the meaning of "substantially the same G-C content", in increasing order of preference, each of the oligonucleotides immobilised on the solid support will have a G-C content within or equal to 25%, 20%, 15%, 10%, 10%, 5% and 2% of the average G-C content of all the immobilised oligonucleotides. In a preferred embodiment, 95-100% of all the oligonucleotides immobilised on the support will have a percentage G-C content within 15   10% of the median value and the remainder will preferably be within 25% of the median value. More preferably the percentage G-C content of the oligonucleotides immobilised on the solid support will be within 8% of each other.

In a preferred embodiment the capture portion of all of the oligonucleotides are of the same length and have the same G-C content.

20   Although it is preferred that all of the oligonucleotides on the solid support fall within the ranges for length and G-C content as defined above, it is envisaged that a small number of oligonucleotides preferably less than 1-5%, more preferably less than 2% of the total number of oligonucleotides may fall outside these ranges. The presence of oligonucleotides on the solid support that possess capture sequences that fall outside the preferred sequence 25   composition (i.e. length and G-C content) need not diminish the utility of the generic array particularly if these capture sequences are excluded from being used as capture molecules when the array is in use.

In a further preferred embodiment there at least 50 different oligonucleotides immobilised on the solid support, and in increasing order of preference, there are at least 100, 30   500, 1000, 5000, 10000 or more different oligonucleotides immobilised on the solid support.

In the most preferred embodiment, there are between 50 and 500 different oligonucleotides immobilised on the solid support.

In a further embodiment, the oligonucleotides are immobilised on the solid support at a density in the range of, in increasing order of preference,1 to 1000 per cm^2, 500 to 1000 per

5   cm^2, 200 to 500 per cm^2, 1 to 200 per cm^2, 1 to 50 per cm^2, and 1 to 10 per cm^2. Most preferred is about 1 per cm^2. In a particular embodiment each distinct capture oligonucleotide is immobilised to the base of a well of a conventional microtitre plate.

Although it is preferred that the oligonucleotides at each pre-determined location on the solid support are unique, this is not essential. Duplicate, triplicate etc., representation of

10   one, more or all capture oligonucleotides on the solid support (array) may be desired in order to detect replicate values.

The capture oligonucleotides attached to the solid support generally have a hybridisable sequence between 5 and 100 nucleotides in length. The preferred length of hybridisable sequence is in the range of 10 - 50, more preferred is 12-30. The hybridisable

15   sequence is that portion of the capture oligonucleotide that is designed and available for hybrid formation with its complementary sequence. As used herein in reference to the capture oligonucleotides, hybridisable sequence and capture portion are used interchangeably. Non-hybridisable sequence of the capture oligonucleotide might represent flanking or tether sequences. Tether sequences not only serve to anchor the oligonucleotide to the solid support

20   but also serve to distance the hybridisable portion of the capture oligonucleotide from the solid support to alleviate steric interference.

The primary structure of each unique capture oligonucleotide on the array can either be designed manually, or can be designed using a computer program to generate random nucleotide sequences. A suitable macro for randomly designing oligonucleotides is disclosed

25   in example 2 herein. It is preferable that none of the capture oligonucleotides on the array (solid support) are capable of hybridising, under stringent hybridisation conditions adopted, to any of the sample target sequences. In this respect, it is preferred that none of the capture oligonucleotides are capable of hybridising with any part of the genome of the test organism under study, be this human, simian, bacterial, viral or the like. In a more preferred

30   embodiment all of the capture oligonucleotides on the array are artificial and lack complementarity to any known sequence from whatever origin. It is also preferred that none

of the capture oligonucleotides cross-hybridise to any test sample nucleic acid. However, if one or more of the capture oligonucleotides on the array do bind to a test sample nucleic acid, this is not detrimental provided that it is known in advance so that any false positive result can be discounted.

5   The capture oligonucleotide molecules may be individually synthesised on a standard oligonucleotide synthesiser. These oligonucleotide (oligos) may then be attached to the substrate matrix by any of a variety of techniques known in the art such as by using photochemical reagents, such as disclosed in US Patent No. 4,542,102 and 4,713,326. US Patent No. 4,562,157 also describes a method of using photo-activatable cross-linking groups

10   to immobilise pre-synthesised ligands on surfaces. Alternatively, the oligonucleotides can be synthesised directly onto the solid surface using photolithography techniques, such as disclosed in US Patent No. 5,143,854, or other methods such as disclosed in US Patent No. 5,700,637, or International Publication No's: WO 95/35505, WO 97/44134 or WO 98/10858. Schena et al. (TIBTECH 16(7):301-306, 1998) reviews the recent advances in microarray

15   technology including the various means of constructing these arrays. Problems facing current photolithographic techniques for oligonucleotide synthesis involve the low yield of synthesis at each synthesis step, and also the efficiency of nucleotide addition at each synthesis step which can range from about 80% to 97%, with purines generally having a lower efficiency than pyrimidines (Thomas & Burke. Exp. Opin. Ther. Patents 8(5):503-508, 1998). When

20   constructing an array with relatively few capture oligonucleotides, say 500 or less, or with long oligonucleotides say 30-mers or more it may be preferable to synthesise the oligos separately and affix them to the solid support later rather than in situ synthesis.

According to a further aspect of the present invention there is provided a method for the preparation of a generic oligonucleotide microarray of the invention, comprising

25   synthesising a plurality of different oligonucleotides and then affixing them to a solid support at a pre-defined location, wherein each oligonucleotide possesses substantially the same T_m as the other oligonucleotides when annealed to its complementary sequence. In a preferred embodiment the oligonucleotides are synthesised on a standard oligonucleotide synthesiser such as an Applied Biosystems model 340A synthesiser.

30   According to a further aspect of the invention there is provided a method for the preparation of a generic oligonucleotide microarray, comprising directly synthesising onto a

solid support at pre-defined positions a plurality of different oligonucleotides, each oligonucleotide possessing substantially the same T_m as the other oligonucleotides when annealed to its complementary sequence. In a preferred embodiment said synthesis is by the photolithography technique as described in US Patent No. 5,143,854.

5   In order to avoid or alleviate steric factors during the capture hybridisation reaction, it may be desirable to use a tether/linker molecule to tether the capture oligonucleotides to the solid support. Shchepinov et al. (N.A.R. 25:1155-1161, 1997) disclose the use of various amino group-containing phosphoramidite moieties to distance the capture oligonucleotide from the solid support and thus alleviate steric interference. They found that with a linker of

10 at least 40 atoms in length they obtained up to 150-fold increased hybridisation yields. Based on the teaching in Shchepinov et al., the person skilled in the art would be able to design and synthesise suitable tether/linker molecules to reduce steric interference of the support on hybridisation behaviour of the immobilised capture oligonucleotides of the invention.

        Thus, in a preferred embodiment the oligonucleotides are attached to the solid support

15 via a tether molecule, such as disclosed in Shchepinov et al. (N.A.R. 25:1155-1161, 1997).

        The novel array with its population of unique capture oligonucleotides is generally used in conjunction with targeting polynucleotide molecules (herein referred to as "targeting polynucleotides").

        The targeting polynucleotides are comprised of two adjacent oligonucleotide

20 sequences, optionally separated by a spacer molecule. The first sequence of the targeting polynucleotide is complementary to one of the capture oligonucleotide sequences on the array. The second sequence is complementary to or substantially complementary to the target sequence to be detected and can therefore act as a detection probe or as an amplification primer. Although the targeting sequence (the second sequence) need not reflect (be precisely

25 complementary to) the exact sequence of the target, the more closely it does reflect the exact sequence the better the binding during the annealing process.

        The term "polynucleotide" as used herein is used to define a molecule composed of ten or more deoxyribonucleotides or ribonucleotides, preferably more than 25. A polynucleotide molecule may be made up of two or more oligonucleotide molecules.

30   The term "complementary to" is used herein in relation to nucleotides to mean a nucleotide which will base pair with another specific nucleotide. Thus adenosine triphosphate

is complementary to uridine triphosphate or thymidine triphosphate and guanosine triphosphate is complementary to cytidine triphosphate. It is appreciated that whilst thymidine triphosphate and guanosine triphosphate may base pair under certain circumstances they are not regarded as complementary for the purposes of this specification. It will also be

5   appreciated that whilst cytosine triphosphate and adenosine triphosphate may base pair under certain circumstances they are not regarded as complementary for the purposes of this specification. The same applies to cytosine triphosphate and uracil triphosphate.

"Precise complementarity" or "perfectly matched" as used herein, is in reference to the duplex that the poly- or oligonucleotide strands make with one another to form a double

10 stranded structure such that every nucleotide in each strand undergoes Watson-Crick base pairing with a nucleotide on the other strand. The term also encompasses the pairing of nucleoside analogues, such as deoxinosine, nucleotides with 2-aminopurine bases, and the like, that may be employed. Conversely, a mismatch in a duplex fails to undergo Watson-Crick bonding.

15 "Substantially complementary" as used herein, refers to poly- or oligonucleotide molecules (or strands) that, under suitable hybridisation conditions, have sufficient complementarity to specifically anneal together, i.e. to the exclusion of all other strands, to form a double stranded structure, but wherein one or other strand has, relative to its partner, a limited number of non-complementary (mismatched) nucleotides that are incapable of

20 undergoing Watson-Crick base pairing with the corresponding nucleotide on the other (partner) strand.

The targeting polynucleotide molecules may be individually synthesised on a standard oligonucleotide synthesiser.

When the targeting polynucleotides are for use as amplification primers, the 5' portion

25 is preferably blocked from acting as template for the polymerase enzyme. This blocking can be effected by linking the 3' portion and the 5' portion of the polynucleotide in the opposite sense to one another, the 3' end being generally in the 5' $\rightarrow$ 3' sense and the 5' end of the polynucleotide being in the 3' $\rightarrow$ 5' sense, with linkage via their 5' termini. The presence of the nucleotide sequence of the 5' portion in the opposite orientation prevents the polymerase

30 enzyme from making a fully double stranded amplification product. The 5' portion of the polynucleotide thus becomes a single stranded tail on the amplification product. This single

stranded tail (the 5' end portion) can then be utilised for capture of the amplification product onto the complementary capture oligonucleotide attached to a solid support. The more preferred means of blocking the polymerisation agent however, is to incorporate a blocking moiety (as the spacer moiety) between the 5' portion and 3' portion of the targeting

5 polynucleotide.

The term " blocking moiety" as used herein means any moiety which when linked, for example covalently linked, between the 3' portion and 5' portion of the polynucleotide is effective to inhibit and preferably prevent, more preferably completely prevent amplification (which term includes any detectable copying) beyond the polymerisation blocking moiety,

10 thus leaving the amplification product with a single stranded tail which is the 5' portion of the polynucleotide. A wide range of blocking moieties may be envisaged for this purpose. For example the polymerisation blocking moiety may comprise a bead, for example a polystyrene, glass or polyacrylamide bead or the polymerisation blocking moiety may comprise a transition metal such as for example iron, chromium, cobalt or nickel (for example in the form

15 of a transition metal complex with the oligonucleotide tail and the target binding nucleotide moiety) or an element capable of substituting phosphorus such as for example arsenic, antimony or bismuth linked between the oligonucleotide tail and the target binding nucleotide moiety. The blocking moiety might similarly involve substitution of the usual phosphate linking groups, for example where oxygen is replaced, leading to *inter alia*

20 phosphorodithioates, phosphorothioates, methylphosphonates, phosphoramidates such as phosphormorpholidates, or other residues known *per se*. Alternative blocking moieties include any 3'-deoxynucleotide not recognised by restriction endonucleases and seco nucleotides which have no 2'-3' bond in the sugar ring and are also not recognised by restriction endonucleases. Newton et al. (Nucleic Acids Research. 21(5):1155-1162, 1993)

25 and EP-B-416817, describe tailed primers with blocking moieties, interposed between the tail and the target binding portion of the primer, that can be incorporated into the polynucleotides of this invention.

In a preferred aspect the spacer comprises a non-amplifiable blocking moiety such as hexethylene glycol (HEG) monomer, alone or combined with further nucleotides, more

30 preferably alone. Alternatively the spacer could comprise material such as 2'-O-alkyl RNA

which will not permit replication of a complementary strand by DNA polymerase enzymes that lack a reverse transcriptase function.

To avoid false positive detection using the microarray and the targeting polynucleotides of the invention, it is desirable that none of the targeting polynucleotides are
5   capable of hybridising to each other. Naturally, it is also desirable that none of the tails (the 5' portions of the targeting polynucleotides) nor spacer moieties are capable of binding to any nucleic acid in the nucleic acid sample. If the capture hybridisation is to be effected in the presence of all the test sample nucleic acid, i.e. without separation of targeting polynucleotide bound nucleic acid from unbound nucleic acid, it is also desirable that none of the capture
10 oligonucleotides on the solid support are capable of binding to any target nucleic acid in the original test sample.

According to a further aspect of the invention there is provided a plurality of polynucleotides, each polynucleotide comprising a unique 3' portion substantially complementary to a unique target nucleic acid sequence which may be present in a sample, a
15 5' portion complementary to one of a group of pre-selected oligonucleotides that each possess substantially the same melting temperature (T_m) and are attached at pre-defined positions to a solid support, and optionally a spacer moiety interposed between said 3' portion and said 5' portion.

In a preferred embodiment, each of the unique 3' portions is precisely complementary
20 with its cognate target sequence. It is to be expected however, that not all the target sequences, complementary to each and every unique 3' portion, will be present in a test sample.

In another preferred embodiment, each of the unique 5' portions from within the population of polynucleotides has substantially the same G-C content and substantially the
25 same length as each of the other 5' end portions.

In a more preferred embodiment, each of the unique 5' portions from within the population of polynucleotides has the same G-C content and length as each of the other 5' end portions.

The 3' portion of the targeting polynucleotide represents the target binding sequence.
30 This sequence can be of any length although it will preferably be between 8 and 60 nucleotides in length, more preferably between 12 and 35 nucleotides in length. The target

binding portion of the targeting polynucleotide sequence need not possess precise complementarity to the target sequence however, it must have sufficient complementarity (be substantially complementary) to bind specifically to the target sequence, that is to say under appropriate hybridisation stringency conditions the target binding region of the primer will

5   hybridise to the target region (if present in the sample) to the exclusion of other regions. There are applications, such as with the ARMS technique, where nucleotide mismatches are incorporated into the target binding primer in order to assist in destabilising primer binding to incorrect target sequences. The presence of certain mismatches need not however, prevent primer binding to the desired target template sequence. In general however, and particularly

10   when relying on allele specific hybridisation, it is preferred that the target binding portion of the targeting polynucleotides has precise complementarity to its target sequence.

The expression "target nucleotide sequence" or "target nucleic acid" or "target sequence" as used herein means a nucleotide or nucleic acid sequence comprising the sequence to be detected by probe or amplified by primer. Thus for example, if the present

15   invention is applied to the diagnosis of β-thalassaemias a sample may contain as many as 60, for example 50, separate potential variant sequences. Each variant sequence is a potential target sequence for probe hybridisation or primer amplification according to the invention disclosed herein.

Amplification of the target sequence can be effected by primer extension off one

20   primer, however, in a preferred embodiment, each targeting oligonucleotide primer is accompanied by a companion primer which facilitates amplification of the target sequence interposed between the two primers according to amplification procedures such as polymerase chain reaction (PCR) or ligase chain reaction (LCR), well known in the art.

When the targeting polynucleotides are for use as allele-specific probes for detecting

25   the presence of a target sequence, the spacer moiety interposed between the two hybridisable elements can serve to prevent steric interference of the two hybrid elements, these two elements being the hybrid duplex molecule consisting of the oligonucleotide targeting portion (the 3' portion) and its complementary sequence from the test sample, and the hybrid duplex consisting of the capture oligonucleotide on the array and the complementary sequence on the

30   targeting oligonucleotide (the 5' portion). The spacer moiety might consist of straight chain or branched alkyl groups, polyglycol residues of any desired number of repeating unit, or

modified nucleotides such as 2'-deoxyribose or 1'napthalene-2'-deoxyribose may be interposed between the first and second portions of the targeting oligonucleotides so as to provide spatial distance between the capture hybrid and the target hybrid.

The capture hybrid refers to the duplex molecule formed by annealing the capture
5 oligonucleotide to the 5' portion (the single stranded "tail") of the targeting polynucleotide. This targeting polynucleotide may or may not already have bound its target sequence. The target hybrid refers to the duplex molecule formed by annealing of the target binding portion (the 3' end) of the targeting polynucleotide to its target sequence. Although the target and capture hybrids have been referred to as duplex molecules, it will be apparent that each
10 complex may not be entirely double stranded.

An advantage of the invention is that the target product, either probe-target hybrid or amplified product, has a single-stranded portion which can be hybridised without denaturation to the solid support containing the immobilise pre-selected capture oligonucleotide sequences. Current array technology requires the target nucleic acids to be denatured or rendered single-
15 stranded some other way, prior to capture on the array.

The microarray and targeting polynucleotides of the invention are useful in any setting where it is desirable to identify the presence of one or more specific nucleic acid sequences from a population of sequences. Examples of uses are in *de novo* or re-sequencing methods, gene expression studies, fingerprinting, diagnostic identification, genotyping of organisms and
20 environmental monitoring.

Any population of nucleic acids represents a suitable test sample. Sources of test sample nucleic acid include human cells such as circulating blood, buccal epithelial cells, cultured cells and tumour cells. Other mammalian tissues and cultured cells are also suitable sources of template nucleic acids. In addition, viruses, bacteriophage, bacteria, fungi and
25 other micro-organisms can be the source of nucleic acid for analysis. The DNA may be genomic or it may be cloned in plasmids, bacteriophage, bacterial artificial chromosomes (BACs), yeast artificial chromosomes (YACs) or other vectors. RNA may be isolated directly from the relevant cells or it may be produced by *in vitro* priming from a suitable RNA promoter or by *in vitro* transcription.

30 The present invention may be used for the detection of variation in genomic DNA whether human, animal or other. It finds particular use in the analysis of inherited or acquired

diseases or disorders. A particular use is in the detection of inherited disease. It will be appreciated that the target nucleic acid is directly or indirectly linked to the sequence or region of interest for analysis.

According to a further aspect of the invention there is provided a method for

5 identifying the presence or absence of one or more test nucleic acid sequences in a sample, comprising the following steps:

i) contacting a nucleic acid containing sample with a plurality of single stranded targeting polynucleotide molecules under suitable hybridisation conditions to ensure hybrid formation between the targeting nucleotide portion of the targeting polynucleotide molecule

10 and its complementary target nucleic acid sequence in the sample, each of said targeting polynucleotide molecules possessing, in addition to the targeting nucleotide portion, a unique single-stranded oligonucleotide tail sequence complementary to a unique capture oligonucleotide attached to a solid support;

ii) contacting the population of hybrid molecules produced in step (i) to a solid support

15 having attached thereon at pre-defined locations unique capture oligonucleotides, each capture oligonucleotide being complementary to one or other of the oligonucleotide tail sequences on the targeting molecules, under suitable conditions to ensure capture of each of the hybrid products to the solid support; and

iii) determining the presence or absence of the captured product at each of the pre-

20 defined locations on the solid support by measurement of a detectable label suitable to identify the captured products;
characterised in that substantially all of the oligonucleotide tails of the targeting polynucleotide molecules and the complementary sequences on the capture oligonucleotides possess substantially the same T_m.

25 In a preferred embodiment, the hybrid molecules produced in step (i) are separated from the unhybridised targeting molecules prior to step (ii). This may be conveniently done by, for example, ethanol precipitation, column chromatography or gel filtration.

In operation, the targeting polynucleotide binds to the target nucleic acid in the sample ("the wet reaction"). If the targeting polynucleotide is serving as a probe, for example as in an

30 allele specific hybridisation, the hybrid molecule is available for capture via the unhybridised

single stranded tail portion of the targeting polynucleotide which is complementary to a capture oligonucleotide attached at pre-defined position on the solid support.

If the targeting polynucleotide is serving as an amplification primer, an amplification reaction must be carried out on the polynucleotide-bound target nucleic acid sample. Multiple

5 rounds of primer extension from the primer can be effected in order to amplify the target nucleic acid such as up to 5, up to 10, up to 15, up to 20, up to 30, up to 40, up to 50 or more times. Conveniently, the targeting polynucleotide is serving as an amplification primer in an amplification system such as the polymerase chain reaction (PCR). In which case the target binding region (3' end) and the tail region (5' end) are advantageously arranged such that the

10 tail region is non-amplifiable in the PCR amplification reaction but remains single stranded, thus ensuring that the amplified product has at least one single stranded tail complementary to a capture oligonucleotide on the solid. This facet of primer design is described in European Patent No. 0 416 817 and corresponding US Patent No. 5525494. In order to effect amplification using PCR-based or LCR-based procedures, a second primer is required. This

15 second oligonucleotide primer may also have a non-amplifiable single stranded tail, possibly identical to the capture portion of the targeting polynucleotide so as to facilitate capture on the solid support. This second primer might also be suitably labelled to facilitate detection of the captured amplification product on the solid support. Alternatively, a suitable label (such as a labelled nucleoside tri-phosphate) might be incorporated into the amplified product during the

20 amplification process.

Any convenient template dependent polymerase may be used, this is preferably a thermostable polymerase enzyme such as Taq™, more preferably Taq Gold™.

Similarly any convenient nucleoside triphosphates for conventional base pairing may be used. If required these may be modified for fluorescence. As these may affect

25 polymerisation rates, for best results, the fluorescently labelled dNTPs are admixed with an excess of wild-type dNTPs, for example in an admixture of between about 1:3 and 1:20.

Further details of convenient polymerases, nucleoside triphosphates, other PCR reagents, primer design, instruments and consumables are given in "PCR" by C.R. Newton and A. Graham (The Introduction to Biotechniques series, Second Edition 1997, ISBN 1

30 85996 011 1, Bios Scientific Publishers Limited, Oxford). Further guidance may be found in

"Laboratory protocols for mutation detection" edited by Ulf Landegren, published by the Oxford University Press, Oxford, 1996, ISBN 0 19 857795 8.

In a preferred embodiment, the targeting polynucleotide molecules are amplification primers, such as those disclosed in EP-B-416817. In a further preferred embodiment, the

5 targeting amplification primer molecules are used in conjunction with another amplification primer to amplify the target region of interest. In a more preferred embodiment, the amplification primers are amplification refractory mutation system (ARMS) primers, as described in EP-B-0332435 and corresponding US Patent No. 5595890, and additionally described in Newton et al. (Nucleic Acids Research. 17(7):2503-2516, 1989). ARMS is a

10 technique suitable for detecting the presence or absence of variant nucleotides at a particular loci. It is particularly useful therefore, in diagnostic detection of mutated nucleic acid indicative of tumour phenotype, or in the detection of single nucleotide polymorphisms (SNPs). ARMS is a particularly useful technique where the target sequence is present in low copy number or there is a need to discriminate between two or more alleles, as for example in

15 mutation detection. ARMS mutation detection enables the sensitive detection of specific alleles in the presence of an excess of alternate alleles. In this way somatic mutations can be detected in a background of wild type DNA. ARMS can readily be used to detect 1% mutant sequence in a 99% wild-type background. ARMS uses primers that allow amplification in an allele specific manner. Allele specificity is provided by the complementarity of the 3'-

20 terminal base of a primer with its' respective allele. Amplification is inhibited when the 3'-terminal base of the primer is mismatched. This specificity is maintained when Taq DNA polymerase or other suitable enzyme lacking 3' to 5' proof-reading activity (such as Klenow) is used. An ARMS test is specific when the yield of product from the target allele exceeds the threshold of detection of the system in use and the yield of product from the nontarget allele is

25 not detectable. As disclosed in EP-B-0332435 the ARMS primers will preferably possess destabilising mismatches incorporated close to the 3'-terminal nucleotide that discriminates between the different alleles, to enhance specific binding and template amplification from the desired allele target sequence. The nearer to the 3' terminus of the primer that a destabilising mismatch is incorporated, the greater the effect on destabilisation (See also Newton et al.

30 Nucleic Acids Research. 17:2503-2516, 1989).

In another preferred embodiment, the second amplification primer for amplifying any particular target sequence possess an oligonucleotide tail identical to that of the first targeting primer. In a further preferred embodiment, the second amplification primer possesses a detectable label, such as a fluorophor or radioisotope to enable the eventual detection of the

5 amplified product on the microarray. Suitable labelling molecules are well known in the art. Alternatively, a suitable label can be incorporated into the amplified product during its synthesis. A suitable amplification reaction can then be performed so as to generate an amplification product.

It will be apparent to the person skilled in the art that there are a large number of

10 analytical procedures which may be used to detect the presence or absence of variant nucleotides at one or more polymorphic positions. Most of these rely on probe or primer hybridisations and thus, with addition of a suitable tail portion to enable capture on an array could be adopted for use with the microarray and method of the present invention. In general, the detection of allelic variation requires a mutation discrimination technique, optionally an

15 amplification reaction and a signal generation system. Table 1 lists a number of mutation detection techniques, some based on the polymerase chain reaction (PCR). These may be used in combination with a number of signal generation systems, a selection of which is listed in Table 2. Many current methods for the detection of allelic variation are reviewed by Nollau *et al.*, Clin. Chem. **43**, 1114-1120, 1997; and in standard textbooks, for example "Laboratory

20 Protocols for Mutation Detection", Ed. by U. Landegren, Oxford University Press, 1996 and "PCR", 2nd Edition by Newton & Graham, BIOS Scientific Publishers Limited, 1997. PCR is described in United States patents nos. 4,683,195 and 4,683,202.

Abbreviations:

ALEX™	Amplification refractory mutation system linear extension
APEX	Arrayed primer extension
ARMS™	Amplification refractory mutation system
b-DNA	Branched DNA
CMC	Chemical mismatch cleavage
COPS	Competitive oligonucleotide priming system
FRET	Fluorescence resonance energy transfer

LCR	Ligase chain reaction
MASDA	Multiple allele specific diagnostic assay
NASBA	Nucleic acid sequence based amplification
OLA	Oligonucleotide ligation assay
PCR	Polymerase chain reaction
SDA	Strand displacement amplification
SSR	Self sustained replication

Table 1 - Mutation Detection Techniques

MASDA; Taqman™ - US-5210015 & US-5487972; Molecular Beacons - Tyagi *et al.* Nature

Biotechnology. 14:303, (1996) and WO 95/13399; ARMS™; ALEX™ - European Patent No.

5 EP 332435 B1; COPS - Gibbs *et al.* Nucleic Acids Research. 17:2347, 1989; APEX; OLA;

SSR; NASBA; LCR; SDA; b-DNA; and minisequencing- Pastinen et al. Genome Research.

7:606-614, 1997.

Table 2 - Signal Generation or Detection Systems

10 **Fluorescence:** FRET, Fluorescence quenching, Fluorescence polarisation - United Kingdom

Patent No. 2228998.

Other: Chemiluminescence, Electrochemiluminescence, Raman, Radioactivity, Colorimetric,

Hybridisation protection assay, Mass spectrometry.

When the targeting oligonucleotides are operating as allele-specific probes, the target

15 nucleic acids is preferably labelled in order to be able to discriminate targeting

polynucleotides bound to the capture oligonucleotides on the array that have target nucleic

acid attached from those that do not. According to one way to do this, the nucleic acid are

degraded to form fragments, degradation is preferably random using for example sonication or

shearing, to generate average lengths of target nucleic acid around the lengths of the

20 complementary sequences on the targeting oligonucleotides. These fragments can then be

labelled. Any number of conventional detectable markers such as radioisotopes, fluorescent

labels, chemiluminescent compounds, labelled binding proteins, magnetic labels,

spectroscopic markers and linked enzymes might be used. One particular example well

known in the art is end-labelling with 32P. Fluorescent labels are preferred because they are

less hazardous than radiolabels, they provide a strong signal with low background and various different fluorophors capable of absorbing light at different wavelengths and/or giving off different colour signals exist to enable comparative analysis in the same analysis. For example, fluorescein gives off a green colour, rhodamine gives off a red colour and both

5 together give off a yellow colour. If the target bound targeting oligonucleotides are not separated from unbound targeting oligonucleotides (following step (i) of the method disclosed herein), and the target nucleic acid is not specifically labelled, other means of discriminating between those captured targeting oligonucleotides that have bound their cognate target sequence from the test sample from those that have not bound their test sample will be

10 required. Suitable means for doing this include the use of intercalating agents (i.e. dyes such as ethidium bromide) that become incorporated into duplex nucleic acid or the use of labelled binding proteins or antibodies or other reagents that recognise helix formation (such as the target nucleic acid/targeting oligonucleotide hybrid), see,e.g., US Patent No. 4,582,789. Convenient intercalators will be apparent to the person skilled in the art (cf Higuchi et al.

15 BioTechnology. 10:413-417, 1992).

It will be apparent to the person skilled in the art that there are other conventional detection means that can be employed in order to detect target bound polynucleotides captured on the solid support. The essential feature is that hybridisation of the captured target nucleic acid bound targeting polynucleotide onto the capture oligonucleotide on the microarray causes

20 a detectable change in a signalling system. Any convenient signalling system may be used, by way of non-limiting example we refer to the measurement of the change in fluorescence polarisation of a fluorescently labelled species (European Patent No. 0 382 433), DNA binding proteins, intercalators, or the incorporation of detectable (modified) dNTPs into the primer extension products or the target nucleic acids.

25 Further systems include two-component systems where a signal is created or abolished when the two components are brought into close proximity with one another. Alternatively, a signal is created or abolished when the two components are separated following binding of the target binding region.

Both elements of the two component system may be provided on the same or different

30 molecules. By way of example the elements are placed on different molecules, target specific

binding displaces one of the molecules into solution leading to a detectable signal. One of the

components may be attached to the capture oligonucleotide, or the solid support itself

For example, the array could consist of fluorescein labelled oligonucleotides of, for

example 20 residues in length. Prior to addition of the sample, a set of short quencher

5   oligonucleotides (say 10 residues in length) complementary to the array and labelled with

DABCYL (or methyl red) could be added to the array. The short complementary DABCYL

oligonucleotides bind to the corresponding 'address' on the array and the fluorescence of the

fluorescein labels is quenched. The sample is then purified so that unextended primers or

unbound probes are separated from extended or bound products. The bound or amplified

10 products are then added to the microarray and the tail portions which are fully complementary

to the oligonucleotides on the microarray bind to the microarray with displacement of the

quencher oligonucleotides. This results in the microarray oligonucleotides fluorescing as a

result of the binding of the appropriate product (see Figure 1). In this format a fluorescent

signal is produced by separating two species bound to the array surface. One advantage of

15 this format is that it permits quality control of the array. When the fluorescent array is

manufactured it can be scanned in a fluorescent scanner and any defects such as a probe

oligonucleotide which has failed to attach to the surface will be detected as a non or weakly

fluorescent spot on the array. Efficient quenching when the quencher oligonucleotides are

added can also be monitored before the test products are added.

20   Convenient two-component systems may be based on the use of energy transfer, for

example between a fluorophore and a quencher. In a particular aspect of the invention the

detection system comprises a fluorophore/quencher pair. Convenient and preferred

attachment points for energy transfer partners may be determined by routine experimentation.

A number of convenient fluorophore/quencher pairs are detailed in the literature (for example

25 Glazer et al, Current Opinion in Biotechnology. 8:94-102, 1997,) and in catalogues such as

those from Molecular Probes, Glen and Applied Biosystems (ABI). Any fluorescent molecule

is suitable for signalling provided it may be detected on the instrumentation available. Most

preferred are those compatible with the 488 nm laser of the ABI PRISM 7700 (Fluorescein

and Rhodamine derivatives). The quencher must be able to quench the dye in question and

30 this may be via a Fluorescence Resonance Energy Transfer (FRET) mechanism involving a

second, receptor fluorophore, or more preferably via a collisional mechanism involving a non-

fluorogenic quencher such as DABCYL, which is a "Universal" quencher of fluorescence or methyl red. Furthermore it is preferred that the selected fluorophores and quenchers are incorporated, most conveniently via phosphoramidite chemistry, into the capture oligonucleotides and/or targeting polynucleotides and/or second primer required for example,

5   when undertaking PCR-based amplification. FAM, a fluorescein dye with an excitation optimum at ~490nm, is a convenient donor.

A further embodiment consists of a fluorescently labelled DNA array where the array oligonucleotides are labelled with a fluorophore which either does not fluoresce at the irradiation frequency or is only weakly fluorescent at this frequency. The ARMS primers are

10   labelled with a fluorophore which is substantially fluorescent at the irradiation frequency and which forms an energy transfer pair with the fluorophore label on the DNA array. When the purified ARMS products are bound to the array and subjected to irradiation there is energy transfer between the fluorophore on the ARMS product and that on the array. The array fluorophore at the specific binding sites increase substantially in its fluorescent brightness and

15   this may be detected by scanning the array. This is an example of increasing the fluorescent signal on the array by bringing two species close together by hybridisation.

The oligonucleotide microarrays and targeting polynucleotides of the invention can be used for large scale hybridisation assays in numerous applications, including genetic and physical mapping of genomes, gene expression studies, sequencing, fingerprinting and

20   genotype mapping, genetic diagnosis and environmental monitoring. The microarray, targeting polynucleotides and method of the invention are particularly suitable for differential gene expression studies. When utilising the generic array and method of the present invention for assessing expression levels of certain genes, RNA can be isolated from a cell or cell population and labelled, for example by attaching a fluorescent molecule to isolated RNA or

25   by end labelling using T4 polynucleotide kinase. Alternatively, mRNA can be reverse transcribed into cDNA and a suitable label incorporated during cDNA synthesis. Fragmentation of these labelled molecules can then precede hybridisation to the targeting polynucleotides prior to capture on the generic array of the invention. As with conventional differential expression studies using gene chips, different fluorescent labels (for example

30   Cy3(green) or Cy5(red)-labelled deoxyuridine triphosphate) can be used on different test samples (i.e. induced versus un-induced) to enable direct comparison of gene expression

levels in the two samples on the same array. The relative fluorescence intensity of each fluor at each array element (capture location) provides a measurement of the relative abundance of the respective RNA in the two cell populations.

With current differential expression studies utilising oligonucleotide arrays, because of
5 the differences in hybridisation properties between short oligonucleotide probes, each target gene must be represented by several oligonucleotides (typically 20 or more) on the chip. In addition, a partner control oligonucleotide identical to each oligonucleotide, except for one of the central nucleotides, is included on the array to serve as an internal control for hybridisation sensitivity. Thus, whereas cDNA arrays only require each gene to be
10 represented by a single hybridisation partner on the array, with the oligonucleotide arrays, each test gene must be represented by at least 40 distinct oligonucleotides each at a different position on the array.

The use of many, for example 5-80, preferably , 10-20, 15-25, 20-30, 25-35, 35-50 distinct targeting polynucleotides per target gene (each range being a separate and
15 independent embodiment of the invention), each complementary to different regions of a particular target gene, and each having the identical tail sequence for capture by a unique capture oligonucleotide on the array, obviates the need for having many distinct oligonucleotides at different locations on the array. Moreover, because each of the capture oligonucleotides possess the same or substantially the same T_m, approximately equivalent
20 capture onto the solid support is expected. This means that the generic array and method of the invention should be particularly suitable for differential expression studies where quantitative analyses are desired. Quantitative analyses in current differential expression studies are limited because of the different hybridisation stabilities of each capture:target duplex sequence. Some sequences will bind more efficiently under a given set of
25 hybridisation conditions than others, hampering precise quantitative analyses. This situation does not arise with the microarray of the present invention because all the capture oligonucleotides on the array possess the same or substantially the same T_m.

Thus according to a further aspect of the invention there is provided the use of the generic microarray, targeting polynucleotides and method of the invention in determining the
30 expression levels of genes from a sample.

A suitable sample might be a tissue sample, for example a biopsy or bodily fluid, such as blood, sample, or a cell sample, for example epithelial or buccal cells, or a cultured cell or cell line such as a mammalian cell or cell line, or a bacterial or yeast cell. Alternatively, it may be from a whole organism, such as *Arabidopsis thaliana*. Any sample containing a

5 plurality of different genes is suitable.

According to a further aspect of the invention there is provided, a method for identifying the differential expression of each of a plurality of genes in a first cell type with respect to expression of the same genes in a second cell type, said method comprising:

(i) isolating mRNA from each cell type and converting said mRNA into cDNA with

10 incorporation of a different fluorescent label into the newly synthesised cDNA for each cell type;

(ii) optionally, fragmenting said newly synthesised cDNA into appropriate length nucleic acid fragments;

(iii) contacting said labelled nucleic acid with a plurality of targeting polynucleotide

15 molecules under suitable conditions to enable hybridisation between substantially complementary sequences to occur, each polynucleotide molecule comprising a unique 3' portion substantially complementary to a unique target nucleic acid sequence which may be present in a sample, a 5' tail portion complementary to one of a group of pre-selected oligonucleotides that each possess substantially the same melting temperature (T_m) and are

20 attached at pre-defined positions to a solid support, and optionally a spacer moiety interposed between said 3' portion and said 5' portion,

(iv) detectably hybridising the hybridised products from step (iii) to a solid surface on which is immobilised at pre-determined positions a plurality of oligonucleotide sequences complementary to one or other of the 5' tail portions of the targeting polynucleotides; and

25 (v) examining the solid support by fluorescence under fluorescence excitation conditions to detect the bound nucleic acid from each cell type, whereby the amount of labelled nucleic acid from each cell type at each particular location on the solid surface can be detected on the basis of the different fluorescence emission colour produced by the different labels incorporated.

30 In a preferred embodiment for differential expression studies, each target gene is detected by a plurality of targeting polynucleotides that bind at distinct parts of the target

gene. In a more preferred embodiment, there are between 5 and 80, preferably about 10-20, 15-25, 20-30, 25-35, 35-50 distinct targeting oligonucleotides per target gene. Each range being a separate and independent embodiment of the invention. In another preferred embodiment, each of the targeting polynucleotide for a particular gene possesses an identical

5 single stranded oligonucleotide tail portion capable of capturing the target:polynucleotide hybrid molecule onto a solid support at a pre-defined position. Although not essential, it is preferred that the nucleotide sequence of each of the capture portions of the targeting polynucleotides and complementary sequences on the capture oligonucleotides on the solid support have substantially the same T_m. In order to facilitate more accurate quantitative

10 analysis, hybridisation conditions are adopted to ensure maximum annealing of the targeting polynucleotides to their target sequences. This may be effected by pooling together those targeting polynucleotides whose target binding sequence possesses on binding to its target approximately the same T_m as the others in the pool. More preferably, the target binding portion of each targeting polynucleotide in a particular pool is of the same nucleotide length

15 and has the same G-C content as the others in the pool. According to a preferred embodiment therefore, each pool of targeting polynucleotides is hybridised to a portion of the test nucleic acid sample under optimum conditions for hybrid formation with their respective target sequences. The hybridisation reactions from each pool of targeting polynucleotides are then pooled together. Following this, capture of the target bound sequences onto the solid support

20 is effected using hybridisation conditions adapted according to the particular T_m of the capture duplexes.

For whatever application, hybridisation conditions chosen are designed to be as close as possible to the T_m of the duplexes. The concentration of salt in the hybridisation solution used is particularly significant. At 1M NaCl, G:C base pairs are more stable than A:T base

25 pairs. Similarly, double stranded oligonucleotides with a higher G-C content have a higher T_m than those of the same length but with a higher A-T content. If slight differences, i.e. single nucleotide differences, amongst the target nucleic acids need to be distinguished, establishing optimum hybridisation conditions is important, particularly, when the hybridisable length of the oligonucleotides is small (< approximately 30-mers). Where,

30 because of the diverse composition of the target sequences, there is a broad range of T_m, either a less than optimum compromise set of hybridisation conditions could be adopted, or

conditions could be manipulated so as to diminish the T_m dependence on nucleotide

composition by using chaotropic hybridisation solutions. This can be effected, for example,

by incorporation into the hybridisation solution of a tertiary or quaternary amide.

Tetramethylammoniumchloride (TMACl) is particularly suitable when used at concentrations

5   of between 2M and 5.5M. A preferred concentration range being 3M - 4M. Compared to the

presence of 1M NaCl in the hybridisation solution, use of up to 5M TMACl can enhance

hybridisation specificity by up to 40-fold. A preferred means of ensuring maximum hybrid

formation despite there being a range of T_m due to target sequence composition, is to divide

the population of targeting polynucleotides into groups according to their optimum T_m, and

10   then to undertake separate hybridisation reactions using sub-groups of pooled targeting

polynucleotides that are grouped according to the T_m of the targeting polynucleotide:target

sequence hybrid portion. Each hybridisation can then be carried out under optimum

hybridisation conditions for the particular group of targeting polynucleotides. In this manner,

optimum hybridisation conditions can be adopted which will ensure approximately equivalent

15   duplex formation. This may be of particular importance if quantitative analysis is required.

The products from the different hybridisation reactions can then be pooled ready for capture

hybridisation to the solid support (Reaction B, see below). An example of a suitable

hybridisation solution involving oligonucleotides of between 15 and 50 nucleotides is: 3M

TMACl, 0.01M sodium phosphate (pH 6.8), 1mM EDTA(pH 7.6), 0.5% SDS, 100 µg/ml

20   denatured, fragmented salmon sperm DNA and 0.1% dried skimmed milk (i.e. Marvel™).

Use of the generic microarray and targeting polynucleotides of the invention requires

two hybridisation reactions to be undertaken. One (A) involves hybrid formation between the

targeting polynucleotides and the target nucleic acid in the test sample ("the wet reaction").

The other (B) involves hybrid formation between the targeting polynucleotides and the

25   capture oligonucleotides bound to the solid support. Depending on the type of study

undertaken, the hybridisations can be carried out in either order or together, although it is

preferred that the wet reaction A is carried out prior to reaction B.

With differential expression studies it is preferred that hybridisation A be carried out

under conditions that ensure maximal hybridisation of the targeting polynucleotides to target

30   sequence. In general, expression monitoring experiments require long overnight

hybridisations with low stringencies (higher salt concentrations, and lower temperatures) in

order to allow hybrid formation between the target nucleic acids and probes that have different stabilities. This also enhances annealing of low copy number sequences. This "wet reaction" can therefore be carried out first in order to allow maximum annealing. Capture of the hybrid molecules from the "wet reaction" can then be effected using greater hybridisation

5   stringencies (at lower salt concentrations and higher temperatures) over shorter time periods (i.e. 1 - 3 hours). The optimum hybridisation conditions can however, be determined from the expected T_m of the capture duplex molecules.

This invention is concerned with a novel microarray of universal use comprising a plurality of oligonucleotides, each possessing substantially the same T_m, attached at pre-

10   defined positions to a solid support. The design and construction of solid support is well known in the art. Essentially, any conceivable solid substrate may be employed in the invention. A suitable substrate is a material having a rigid or semi-rigid surface, generally insoluble in a solvent of interest such as water. Specific suitable substrates are glass, plastics, polymers, polysaccharides, resins, metal, silica or silica-based material, nylon or

15   nitrocellulose filters, and the like. The solid support may comprise a single sheet of a suitable material such as glass, silicon or plastic so that the pre-selected oligonucleotides are positioned at pre-defined sites based on each oligonucleotide having a distinct and distinguishable set of x,y co-ordinates. Alternatively the solid support may comprise a set of beads of a suitable material such as glass or plastic so that the pre-selected oligonucleotides

20   are positioned at pre-defined sites based on each oligonucleotide residing on a distinct bead. In a preferred embodiment the substrate and/or its surface will be flat glass or single-crystal silicon. Suitable examples of existing laboratory materials that can be utilised are glass microscope slides and microtitre (such as 96-well) plates. The surface of the substrate will preferably contain reactive groups such as carboxyl, amino, hydroxyl, or the like. A

25   polycationic polymer such as polylysine is particularly useful. Most preferably, with fluorescence detection, the surface is non-fluorescent at the wavelength that the analysis is to be performed. The surface of the substrate is also preferably provided with a layer of cross-linking groups to assist attachment of the oligonucleotides to the support. These cross-linking groups will preferably be of sufficient length to permit the oligonucleotides attached to

30   interact freely with their binding partners in solution. Crosslinking groups may be selected from any suitable class of compounds, for example, aryl acetylenes, ethylene glycol oligomers

containing 2-1 monomer units, diamines, diacids, amino acids, and the like. The cross-linking groups may be attached by a variety of methods which are readily apparent to the person skilled in the art. For example, by esterification or amidation reactions of an activated ester of the linking group with a reactive hydroxyl or amine group on the free end of the cross-linking

5 group.

The detection of specific interactions may be performed by detecting the positions where the labelled target sequences are attached to the array. Radiolabelled probes can be detected using conventional autoradiography techniques. Use of scanning autoradiography with a digitised scanner and suitable software for analysing the results is preferred. Where the

10 label is a fluorescent label, the apparatus described, e.g in International Publication No. WO 90/15070, US Patent No. 5, 143,854 or US Patent No. 5,744,305 may be advantageously applied. Indeed, most array formats use fluorescent readouts to detect labelled capture:target duplex formation. Laser confocal fluorescence microscopy is another technique routinely in use (M.J.Kozal et al., Nature Medicine. 2:753-759, 1996). Mass spectrometry may also be

15 used to detect oligonucleotides bound to a DNA array (Little DP et al, Analytical Chemistry. 69(22):4540-4546, 1997). Whatever the reporter system used, sophisticated gadgetry and software may be required in order to interpret large numbers of readouts into meaningful data (such as described, for example, in US Patent No. 5,800,992 or International Publication No. WO 90/04652).

20 Once a particular sequence has, or group of sequences have, been hybridised to the microarray and the pattern of hybridisation analysed, the microarray can be treated to remove the bound sequences in preparation for reuse of the microarray by exposure to a second or subsequent set of target sequences. In order to do this the hybrid duplexes are disrupted and the solid support matrix treated in order to remove all traces of the original target. To effect

25 this, the matrix may be treated with various detergents or solvents to which the substrate, the oligonucleotides and the linkages to the substrate are inert. This treatment may involve an elevated temperature treatment, treatment with organic or inorganic solvents, modifications in pH, and other means for disrupting specific interactions. Examples of methods that could be used are: (1) Washing the array with 50 mM sodium hydroxide to disrupt base pairing by high

30 pH. (2) Washing the array with pure water and at high temperature (e.g. > 80°C) to disrupt base pairing by high stringency. (3) Addition of oligonucleotide sequences complementary to

the tail sequences (and identical to the chip sequences) to disrupt base pairing by exchange with the sequences in free solution. Other methods for disrupting duplex formation are well known in the art (see for example Sambrook et al. *ibid*). Because the microarray of the invention is not a custom chip, but rather a generic chip which interacts with specific custom

5 targeting oligonucleotides, once the microarray has been cleaned, it can be reused in any appropriate procedure and is not limited to reuse in the particular procedure used before. The discriminatory ability lies with the "wet reaction" involving the target nucleic acids with the custom targeting polynucleotides that co-operate with the generic microarray of the invention.

According to a further aspect of the invention there is provided a kit for detecting the

10 presence or absence of one or more target nucleic acid sequences contained in a sample, which kit comprises:-

(1) a plurality of polynucleotides, each polynucleotide comprising a unique 3' portion substantially complementary to a unique target nucleic acid sequence which may be present in a sample, a 5' portion complementary to one of a group of pre-selected oligonucleotides that

15 each possess substantially the same melting temperature (T_m) and are attached at pre-defined positions to a solid support, and optionally a spacer moiety interposed between said 3' portion and said 5' portion; and

(2) a solid support having immobilised thereon a plurality of pre-selected oligonucleotides at pre-defined positionally distinct sites, characterised in that the composition of each of the

20 oligonucleotides is such that they all have substantially the same melting temperature (T_m) when annealed to their complementary sequence, each capture oligonucleotide having a sequence complementary to a 5' portion of one of the primers in (1).

In a preferred embodiment the kit also contains each of four different nucleoside triphosphates and an agent for polymerisation of the nucleoside triphosphates. In another

25 preferred embodiment the kit comprises a set of at least two primers for each target sequence, the terminal nucleotide of at least one primer being complementary to a suspected variant nucleotide associated with a known genetic disorder and at least one of the other primers being a companion primer as described hereinbefore. In a further embodiment, the polynucleotide molecules referred to in (1) are ARMS primers with non-amplifiable tails as

30 described hereinbefore. In a further embodiment the solid support is a microscope slide or

microtitre plate, such as a 96-well plate. Such kits may also comprise control DNA and control primers or probes, and the like.

The invention will now be further illustrated by the following non-limiting examples. The examples refer to the following figures, in which:

5   Figure 1 - Illustrates the use of quenching oligonucleotides for measuring primer binding.

Figure 2a - Results from Experiment 1, 84-ARMS primer multiplex on 1% p53 codon 175 CAC admixture template and wild-type; OD 405nm mutant (ODm) and wild-type (ODw) plot for the 11 test ARMS primers and one control primer.

Figure 2b - Results from Experiment 1, (ODm/ODw)-1 ratio of absorbance (405nm) on p53

10 wild-type and mutant admixture for 1% 175 CAC.

Figure 3a - Results from Experiment 1, 84-ARMS primer multiplex on 5% p53 codon 175 CAC admixture template and wild-type; OD 405nm mutant (ODm) and wild-type (ODw) plot for the 11 test ARMS primers and one control primer.

Figure 3b - Results from Experiment 1, (ODm/ODw)-1 ratio of absorbance (405nm) on p53

15 wild-type and mutant admixture for 5% 175 CAC.

Figure 4a - Results from Experiment 2, 84-ARMS primer multiplex on 1% p53 codon 175 CAC admixture template and wild-type; OD 405nm mutant (ODm) and wild-type (ODw) plot for the 17 test ARMS primers and two control primer.

Figure 4b - Results from Experiment 2, (ODm/ODw)-1 ratio of absorbance (405nm) on p53

20 wild-type and mutant admixture for 10% 175 CAC.

Example 1

Multiplex tailed ARMS assay to detect p53 mutations.

Many potential mutation sites in p53 have been identified (P.Hainaut et al, Nucleic

25 Acids Research. 26(1):205-213, 1998).

80 ARMS primers were designed for the specific detection of some of the mutations in exons 5-8 of the p53 tumour suppresser gene (Table 3 lists the 80 codon positions and specific mutations for which ARMS primers were designed and prepared). Uniquely identifying non-amplifiable tails (with T_ms in the range of 53°C to 58°C) with hexaethylene

30 glycol links between the primer and the tail sequences were added to 19 of these ARMS primers (marked * in Table 3). The 80 ARMS primers were then multiplexed together with 2

reverse primers designed to give PCR products with the ARMS primers specific for mutations in exons 5&6 and 8 respectively. Tailed primer sets which act as control primers for the detection of p53 exons 5&6 and exon 8 sequence were also included.

Table 3.

5 List of potential p53 mutations on which the ARMS primers were prepared.

132 CAG*	132 AGG	135 TAC	141 TAC	151 TCC	151 ACC*
152 CTG	154 GTC	156 CCC*	157 TTC*	158 CTC*	158 CAC
159 GAC*	159 GTC	161 ACC	163 TGC	173 ATG	173 TTG*
175 CAC*	176 TTC*	177 CGC*	179 CGT	179 TAT	192 TAG
193 CGT	195 ACC	196 TGA	205 TGT	213 TGA	220 TGT*
228 AAC	234 TGC	237 ATA	238 TAT	241 TTC	242 TTC
244 TGC	245 TGC	245 GAC	245 AGC	245 CGC	245 GTC
245 GCC	248 CTG	248 CAG	248 TGG	248 GGG	248 CCG
249AGC	249AGT	249 ACG	249 ATG	249 GGG	249 AAG
258 AAA	258 GGA	266 AGA	266 GAA	266 TGA	266 GTA
272 ATG	273 CCT*	273 CTT*	273 CAT*	273 TGT*	273 AGT
273 GGT	275 TAT	278 CTT	278 TCT	280 ACA*	282 TGG
282 CAG	282 CCG	282 CTG	282 GGG	285 AAG*	286 AAA*
298 TAG*	306 TGA				

* - denotes the mutant codon for which a tailed ARMS primer was prepared.

10   Templates containing mutations in the p53 gene were prepared by primer directed mutagenesis (see Higachi et al. NAR. 16:7350-7367, 1988). This gave templates of 500 - 800 bp containing the mutation of interest. Wild type templates were prepared by PCR amplification of wild-type DNA to give the corresponding 500 - 800 bp fragments.

All templates were quantified relatively by real time analysis on an ABI Prism 7700

15 using a quantitative PCR reaction. Quantitated cassettes (mutant synthetic templates prepared by site directed mutagenesis) were then used to prepare wild-type/mutant admixtures.

- 33 -

Oligonucleotides of precise complementary sequence to the 19 tail sequences were synthesised with 3' biotin moieties. These capture sequences were bound to the wells of a streptavidin coated microtitre plate (one capture sequence per well).

Reaction conditions: Each ARMS primer was present at 50 nM concentration, the
5 reverse primers were present at 500mM concentration and wild type dNTPs at 50 μM each. Fluorescein-dUTP was also included at 0.5 μM. The buffer was 50 mM KCl, 10 mM tris, 1.2 mM $MgCl_2$ at pH 8.3. 4 Units of AmpliTaq Gold™ were used per amplification. 10^5 copies of template were added. Cycling conditions were 94°C for 20 minutes then 35 cycles of (94°C for 1 minute, 60°C for 1 minute). In each experiment, three parallel experiments were
10 run using (a) wild type template, (b) mutant/wild-type admixture template and (c) a no-template control.

Following amplification, the PCR products were divided between the capture wells of the microtitre plate. Hybridisation between the PCR products and the capture oligos took place overnight at 55°C in 3M TMAC, 1M Tris (pH 7.5), 0.5M EDTA, 0.01% Triton-X-100,
15 0.1mg/ml herring sperm DNA. Unbound products were then washed off (2 washes in phosphate buffered saline (PBS)). The PCR products were detected by ELISA detection of incorporated fluorescein-dUTP using an anti-fluorescein-alkaline phosphatase antibody-enzyme conjugate. Colour development was by addition of p-nitrophenyl phosphate and the OD 405 was determined after 30 minutes.

20   For each primer being examined the following ODs were obtained: (i) the mutant template termed ODm; (ii) the wild-type template termed ODw; and, (iii) the no-template control. The no-template control ODs were subtracted from ODm and ODw to give background corrected values. ODm, ODw values and the (ODm/ODw -1) ratio were then plotted.

25

Experiment 1 - Detection of a single point mutation at codon position 175 of p53 present as template at a concentration of 1% or 5%.

Admixtures were prepared between template containing p53 mutant codon 175 CAC and wild-type template with the mutant sequence present at 1% and 5% of the total.

30

Specific tailed ARMS primer used to detect p53 175 CAC mutation:

5'-GCTTTATGTCCACAGATTTC*ATACACAGCACATGACGGAGGTTGTGAGCCA-3'

The * denotes the HEG group.

Specific reverse primer used with the above ARMS primer:

5   5'-ACCCGGAGGGCCACTGACAAC-3'

For each admixture three separate amplifications (using the multiplex of 80 mutant
ARMS primers and 2 control primers, 19 of which had tails so that they could be captured-see
Table 3) were carried out on: (a) the admixture; (b) wild-type template; and, (c) no-template

10   control. The amplification products were then each added to a separate array consisting of a
microtitre dish with 11 capture oligos for a subset of the primers plus 1 capture oligo for the
exon 5&6 control reaction immobilised thereon.

The ODm and ODw values with the relevant no-template control values subtracted are
shown in figures 2a and 3a. The figures show that the primer for 175 CAC gives a signal on

15   the mutant template (bars) which is higher than the one obtained with wild-type template
(diamonds). The size of the signal is dependant on the amount of mutant template present in
the substrate and is therefore greater for the 5% mutant template than for the 1% mutant
template.

The ability of the primer for 175 CAC to differentially detect low levels of mutant

20   sequence can be also seen by calculating the (ODm/ODw-1) ratios (see Figures 2b and 3b).
From these values it can be clearly seen that only the primer for 175 CAC demonstrates a
significantly higher OD on the mutant template than on the wild-type template.

Careful examination of the OD405nm data and the (ODm/ODw)-1 plots permits one
to distinguish between primers which are selectively detecting mutant sequence and those that

25   are giving unselective amplification from wild-type as well as mutant template. Experiment 2
illustrates this more fully.

Experiment 2

Detection of a single point mutation at codon position 175 of p53 present as template at a

30   concentration of 10%.

An admixture was prepared between template containing p53 mutant codon 175 CAC at 10% in a wild type background. As in experiment 1, three separate amplifications were carried out on: (a) the admixture; (b) wild-type template; and, (c) no-template control. The products of the amplifications were then each added to a separate array consisting of 17

5   capture oligos for the primers plus 2 capture oligos for exon 5, 6 and exon 8 control reactions. The ODm and ODw values with the relevant no-template control values subtracted are shown in Figure 4a. This figure shows that the primer for 175 CAC has given a signal on the mutant template (bars) which is far higher than the one obtained with wild-type template (diamonds). It can also be seen that the 220 TGT mutation has a propensity to prime and give detectable

10   product with wild-type template as well as on the mutant template. That this primer is not erroneously detecting mutant sequence can be seen by calculating the (ODm/ODw-1) ratios (see Figure 4b) in which case it can be clearly seen that only the primer for 175 CAC is giving a higher OD on the mutant template than on the wild-type template.

15   Summary:

Example 1 demonstrates the use of ARMS for the sensitive detection of under-represented sequences in combination with the use of non-amplifiable tails and oligonucleotide arrays to provide a method for the large scale multiplex analysis of polymorphisms in gene sequences.

20   The use of ARMS in solution phase permits the more sensitive detection of gene variation than can be achieved with solid phase allele specific oligonucleotide (ASO) hybridisation.

The use of non-amplifiable tails and oligonucleotide arrays presents a more generic and widely applicable technique then can be achieved with ASO arrays which must be

25   individually designed on a target to target basis. In this way a maximally efficient mutation detection system is produced because each component used in the process is suited to the process it is required to carry out. Mutation detection can be carried out in solution phase using ARMS and DNA arrays used for separating complex mixtures of oligonucleotide sequences.

30   The use of non-amplifiable tails and oligonucleotide arrays presents a simpler hybridisation technique then can be achieved with ASO arrays. With ASO arrays all of the

probe to targets hybridisations require to be carried out under the same conditions including the same temperature and buffer. These conditions may not be ideal for many of the probe to target hybridisations which require to be performed. With generic arrays and non-amplifiable tails a single set of unified hybridisation conditions can be pre-selected which permit all probe

5   to target hybridisations to be carried out under the same optimal conditions because all probe to target hybridisations can be selected to occur between sequences of substantially the same T_m and GC content.

The use of non-amplifiable tails and oligonucleotide arrays also presents a more cost effective way of screening for mutations than can be achieved with ASO arrays. The

10   manufacture of the array used in the screening process is cheaper than is the case with specific ASO arrays because:

1. Depending on its use, the generic array will likely require far fewer capture sequences for the analysis of each variant in the gene of interest than is possible with a specific ASO array and therefore, is far simpler and cheaper to manufacture.

15   2. Once the array has been designed it may be used for the analysis of any gene target. The costs of designing and developing new arrays for new gene targets is avoided.

3. With one array being used for the analysis of all gene targets economy of scale can be realised during manufacture compared to the situation where smaller manufacturing runs are undertaken to produce a multitude of specific arrays.

20   Example 2

Design of suitable oligonucleotides with substantially the same T_m.

Oligonucleotide sequences of substantially the same T_m can conveniently be generated by use of a spreadsheet computer program incorporating a random number generator. By way of example the following Visual Basic macro, when run in Microsoft Excell™, will generate

25   random sequences of between 10 and thirty bases in length. The T_m of these bases is then calculated using the simple algorithm $T_m = [2* (\#A \text{ or } T) + 4*(\#G \text{ or } C)]$ (i.e. each A or T base pair adds 2°C to the T_m while a G or a C adds 4°C). The program then sorts the sequences in order of increasing T_m. In this way oligos of substantially the same T_m can be selected as candidate sequences for use as tails. It should be understood that the use of this

30   T_m algorithm is illustrative only and that any convenient algorithm could be used.

Macro:

```
Option Explicit
Sub probe()

    'Declare arrays
    Dim well
    Dim number
    Dim repeat
    Dim percent
    'get number of sequences to generate
    Let number = Application.InputBox("Enter number of random sequences to generate")
    'redeclare arrays
    ReDim NextCell(30)
    ReDim Character(30)
    ReDim NoBases(number)
    ReDim NoGC(number)
    ReDim PercentGC(number)
    ReDim Tm(number)
    Dim limit
    Dim total
    'Generate Probe Sequences
    For repeat = 1 To number
    'determine sequence length
    Range("BB1").Select
    ActiveCell.FormulaR1C1 = "=ABS(RAND())"
    Range("BB1").Select
    Let limit = (ActiveCell * 20) + 10
    'generate sequence
    For well = 1 To limit
    Range("A1").Select
```

```
        ActiveCell.Offset(repeat, well).FormulaR1C1 = "=ABS(RAND())"
        Range("A1").Select
        ActiveCell.Offset(repeat, well).Select
        Selection.Copy
5       Range("A1").Select
        ActiveCell.Offset(repeat, well).Select
        Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone, _
            SkipBlanks:=False, Transpose:=False
        Next well
10      'convert to bases, detect G and C, total Tm
        Let NoBases(repeat) = 0
        Let NoGC(repeat) = 0
        Let Tm(repeat) = 0
        For well = 1 To limit
15      Range("A1").Select
        Let NextCell(well) = ActiveCell.Offset(repeat, well)
        If (NextCell(well) < 0.25 And NextCell(well) > 0#) Then
        Let Character(well) = "A"
        Let NoBases(repeat) = NoBases(repeat) + 1
20      Let Tm(repeat) = Tm(repeat) + 2
        End If
        Let NextCell(well) = ActiveCell.Offset(repeat, well)
        If (NextCell(well) < 0.5 And NextCell(well) > 0.25) Then
        Let Character(well) = "C"
25      Let NoBases(repeat) = NoBases(repeat) + 1
        Let NoGC(repeat) = NoGC(repeat) + 1
        Let Tm(repeat) = Tm(repeat) + 4
        End If
        Let NextCell(well) = ActiveCell.Offset(repeat, well)
30      If (NextCell(well) < 0.75 And NextCell(well) > 0.5) Then
        Let Character(well) = "G"
```

```
        Let NoBases(repeat) = NoBases(repeat) + 1

        Let NoGC(repeat) = NoGC(repeat) + 1

        Let Tm(repeat) = Tm(repeat) + 4

        End If

  5     Let NextCell(well) = ActiveCell.Offset(repeat, well)

        If NextCell(well) > 0.75 Then

        Let Character(well) = "T"

        Let NoBases(repeat) = NoBases(repeat) + 1

        Let Tm(repeat) = Tm(repeat) + 2

 10     End If

        Next well

        Range("A1").Select

        For well = 1 To limit

        ActiveCell.Offset(repeat, well).FormulaR1C1 = Character(well)

 15     Next well

        Range("A1").Select

        ActiveCell.Offset(repeat, 0).FormulaR1C1 = repeat

        'calculate %GC

        Let PercentGC(repeat) = (NoGC(repeat) / NoBases(repeat)) * 100

 20     Next repeat

        'output calculations

        Range("AG1").Select

        For repeat = 1 To number

        ActiveCell.Offset(repeat, 0).FormulaR1C1 = NoBases(repeat)

 25     Next repeat

        Range("AH1").Select

        For repeat = 1 To number

        ActiveCell.Offset(repeat, 0).FormulaR1C1 = PercentGC(repeat)

        Next repeat

 30     Range("AI1").Select

        For repeat = 1 To number
```

```
ActiveCell.Offset(repeat, 0).FormulaR1C1 = Tm(repeat)

Next repeat

Columns("A:AF").Select

Selection.ColumnWidth = 4

ActiveWindow.SmallScroll ToRight:=17

Columns("AH:AH").Select

Selection.NumberFormat = "0.0"

Range("AG1").Select

ActiveCell.FormulaR1C1 = "# bases"

Range("AH1").Select

ActiveCell.FormulaR1C1 = "% GC"

Range("AI1").Select

ActiveCell.FormulaR1C1 = "Tm °C"

Range("A1").Select

ActiveWindow.ScrollColumn = 1

Columns("A:AI").Select

Selection.Sort Key1:=Range("AI2"), Order1:=xlAscending, Header:= _
    xlYes, OrderCustom:=1, MatchCase:=False, Orientation:= _
    xlTopToBottom

Range("A1").Select

End Sub
```

The design and choice of suitable tail sequences with substantially the same T_m can of course also be carried out manually using the simple algorithm $T_m = [2* (\#A \text{ or } T) + 4*(\#G \text{ or } C)]$.

Example 3.

A suitable format for differential expression studies.

cDNA is prepared from mRNA purified from two types of tissue representing normal and altered cells. The altered cells may have been treated differently to the normal cells prior to mRNA purification. Examples of 'different treatment' include starving the cells of a

metabolite or metabolites, stimulating them with a specific metabolite such as a growth factor or treating them with a drug or hormone. Alternatively, different cell conditions might exist already, i.e. normal Vs tumour.

5   Typically, reverse transcription from an oligo-dT primer is carried out with incorporation of fluorescent labels into the cDNA prepared. Often one label such as Cy-3 is used to label the cDNA from the normal tissue and another label such as Cy-5 is used to label the cDNA from the treated tissue. The cDNA population is optionally fragmented by sonication or mechanical shearing (i.e. by passage through a 19G needle). Targeting polynucleotides are added to the cDNA mixture and hybridisation permitted to occur, for

10   example, under the following hybridisation condition: labelled cDNA is resuspended in 10 ml of 3.5 x SSC containing 4mg of poly dA DNA, 2.5 mg *E. coli* tRNA, 4mg human Cot1 DNA and 1ml 10 10% SDS. Hybridisation is carried out at 62°C for 3hours.

To facilitate quantitative analysis and ensure efficient target binding and identification, each cDNA is targeted by a number of distinct targeting polynucleotides each capable of

15   hybridising to a different region of the cDNA, but all possessing the same tail sequence to facilitate capture at the same pre-determined location on the microarray. By way of example, each cDNA is targeted by 20 or more distinct targeting polynucleotides. As described above, the targeting polynucleotides comprise two principal domains, the first domain is complementary to one of the cDNAs in the sample mixture and the second domain (tail

20   portion) is complementary to one of the capture oligonucleotides on the DNA array. If the range of T_m between the target sequences and their complementary sequences on the targeting polynucleotides is wide, it may be preferable to group the targeting polynucleotides into pools of substantially the same T_m. In this way, separate sample:targeting polynucleotide hybridisations can be carried out under the optimum hybridisation conditions for each pool.

25   All the pooled reaction products can then be mixed prior to the capture hybridisation. Once hybridisation between the cDNA and the targeting polynucleotides is substantially complete, the mixture is added to the surface of the oligonucleotide microarray under suitable conditions to allow hybridisation between the targeting polynucleotide tail portion to occur. Hybridisation is carried out at 62°C for 1-3hours in a suitable volume of hybridisation

30   solution such as10ml of 6x SSC, 0.1%SDS and 0.25% dried skimmed milk (Marvel™) in a suitable enclosed vessel. A proprietary hybridisation apparatus such as model HB-1 (Techne

Ltd) provides reproducible conditions for the experiment. On completion of hybridisation the microarray is subjected to a stringency wash (such as in 2xSSC, 0.2%SDS, then 0.2xSSC) and the array surface is subjected to fluorescence. Fluorescent output from the two dyes is captured and stored as separate channels. The intensity of the two data sets is normalised by

5 reference to a common housekeeping gene whose expression is considered to be invariant in all tissues. There are many such genes but one example is GAPDH.

Having normalised the data, the differences in intensity at each point on the microarray is measured. Up or down regulation of genes in the treated tissues would be seen as increases or decreases respectively in the intensity of the Cy-5 signal compared to the intensity of the

10 same array spot in the Cy-3 channel.

DNA array with fluorescently labelled capture oligonucleotides:
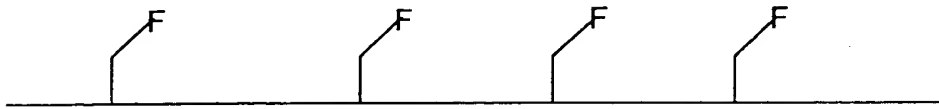
Addition of short quencher oligonucleotides to quench array:
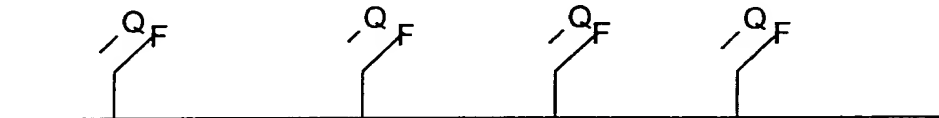
addition of purified amplification products to array:
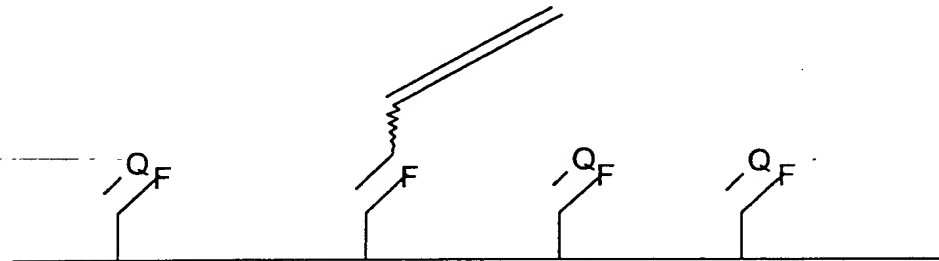
Figure 1. Use of quenching oligonucleotides (Q) for measuring primer binding. F represents fluorescent label attached to capture oligonucleotide.
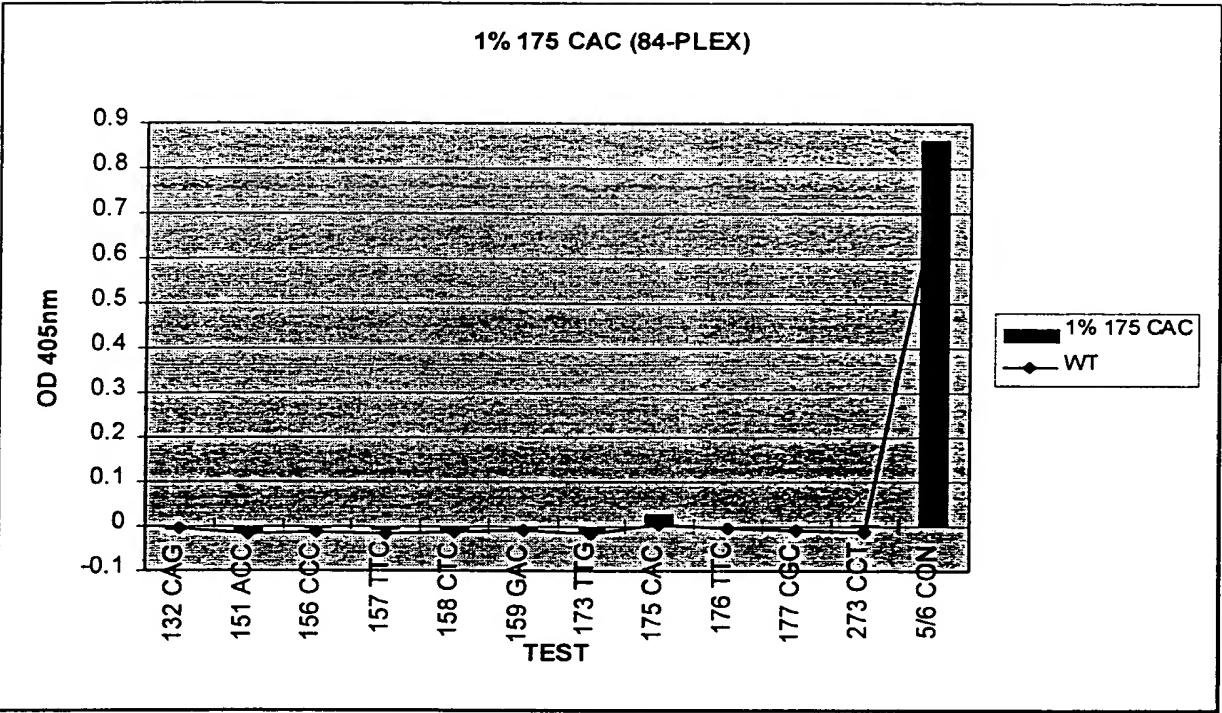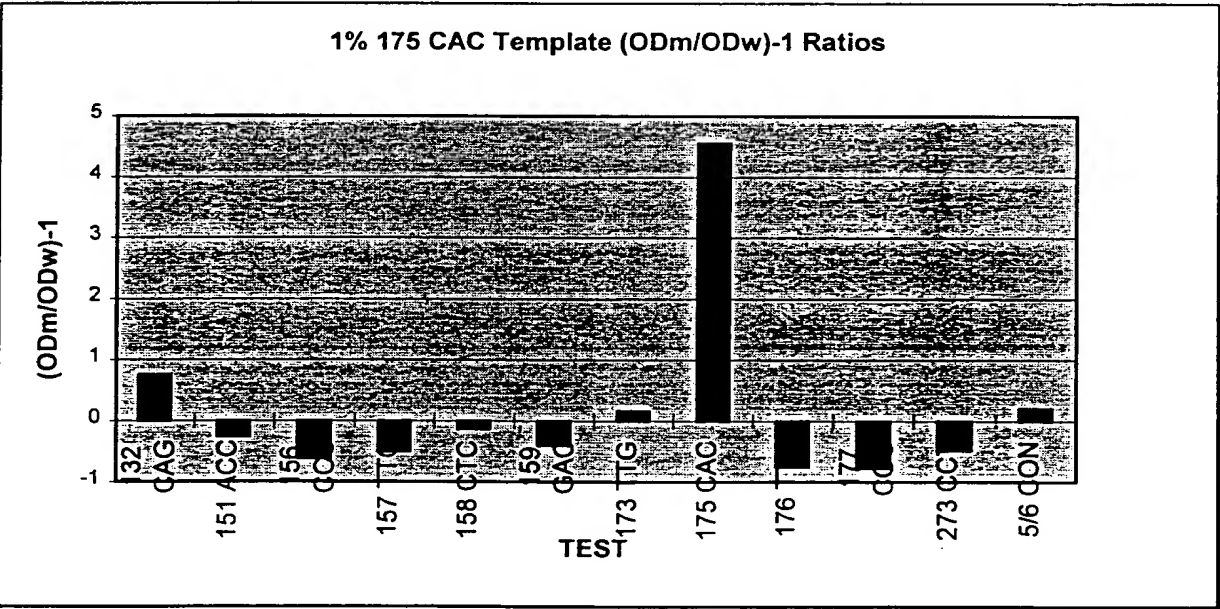
Figure 2a. Experiment 1. ODm (bars) and ODw (line)
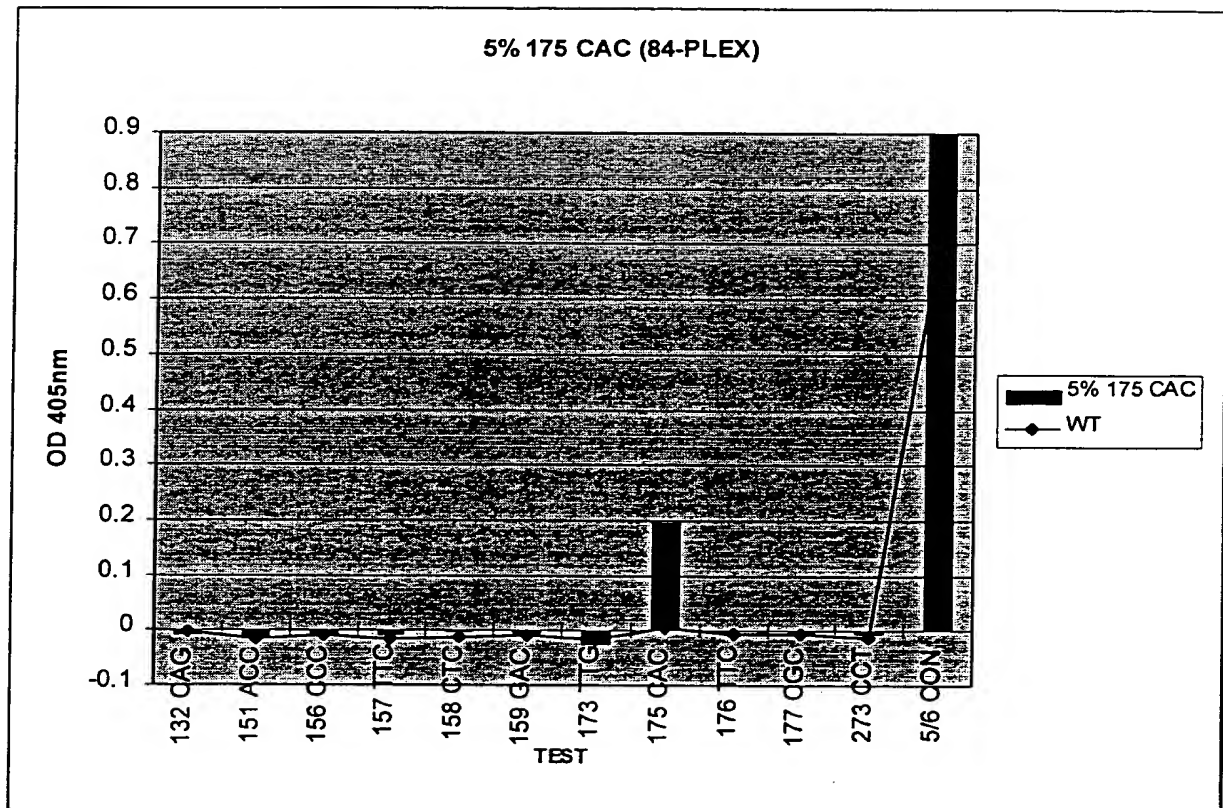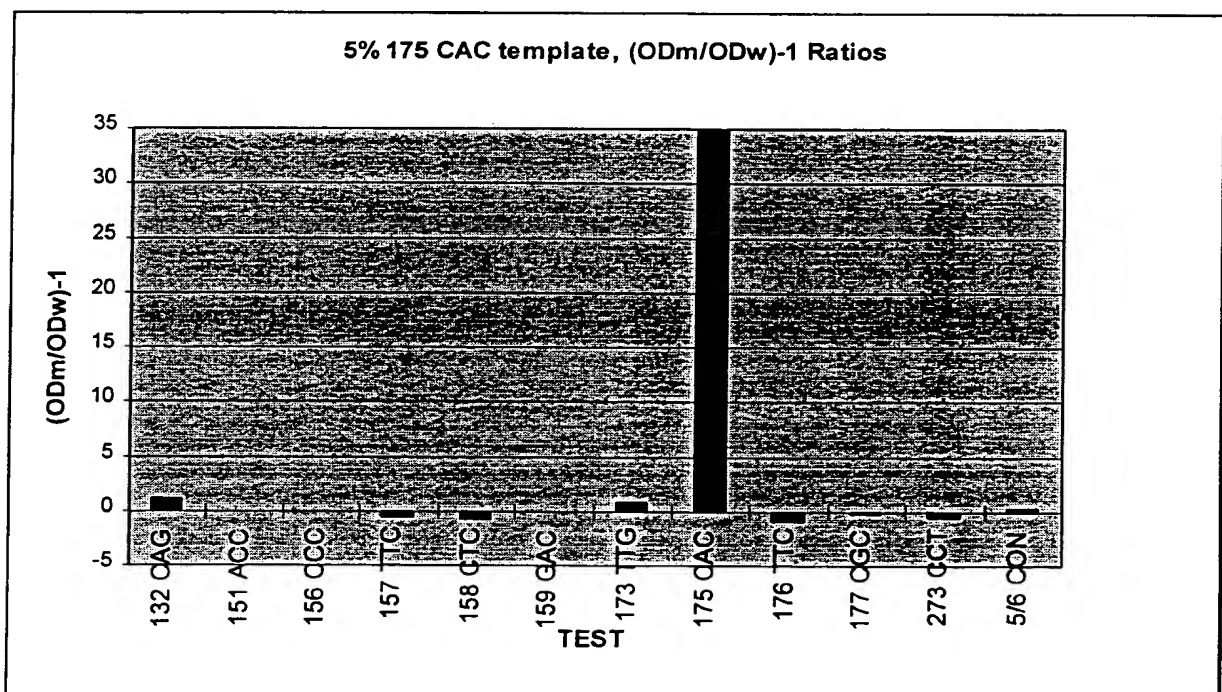

Figure 2b. Experiment 1. (ODm/ODw)-1

Figure 3a. Experiment 1. ODm (bars) and ODw (line)
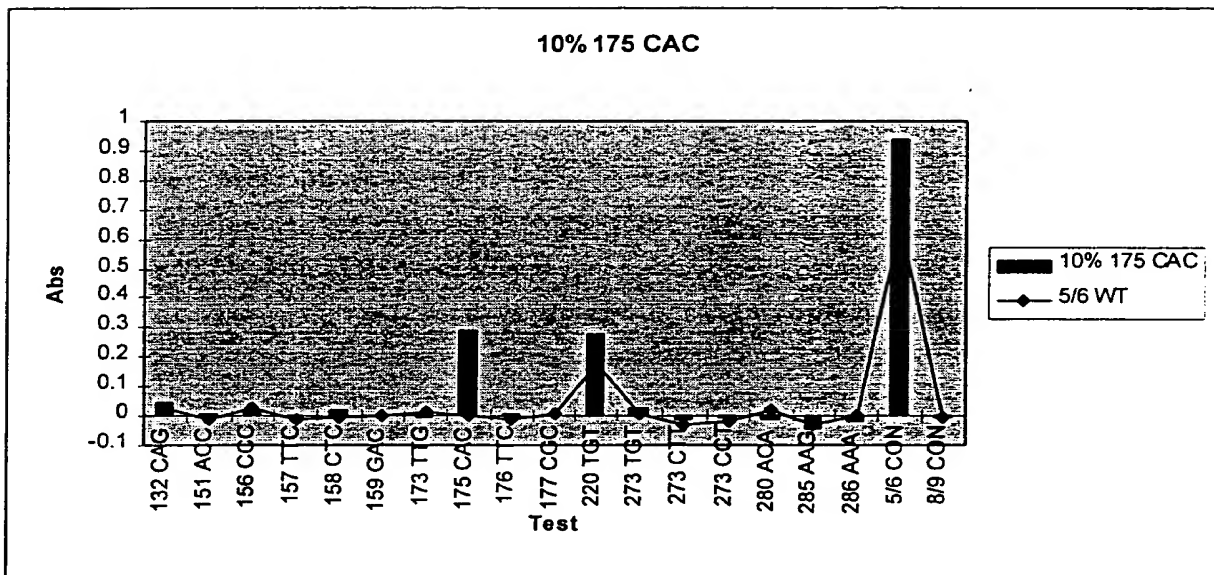

5 Figure 3b Experiment 1. (ODm/ODw)-1
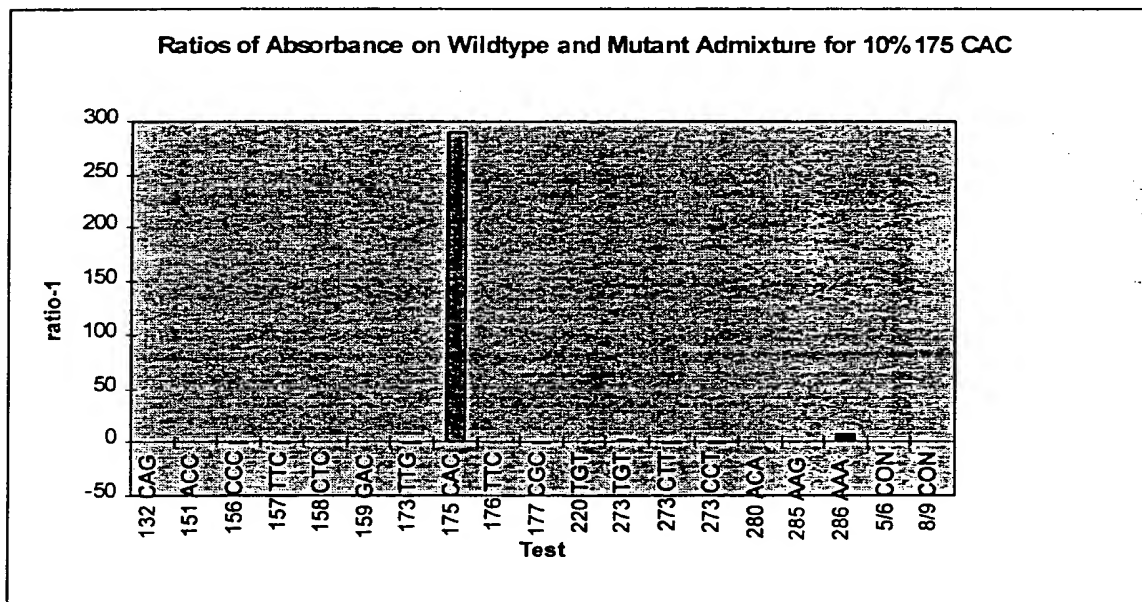
Figure 4a Experiment 2. ODm (bars) and ODw (line)

Figure 4b Experiment 2. (ODm/ODw)-1

PCT/US00/00857

AstraZeneca

8/2/00.